DEEPESH YADAV

Bangalore, KA

https://github.com/deepeshyadav760 www.linkedin.com/in/deepesh-yadav-b29a29250 deepeshyadav760@gmail.com

+91- 7275632538

SUMMARY

Data Science & AI enthusiast with hands-on experience in ML, DL, and Generative AI (RAG, Transformers, LLMs). Skilled in Python, SQL/NoSQL, model deployment, real-time pipelines, and frameworks like PyTorch, TensorFlow, and Scikit-Learn. Experienced in recommender systems, search optimization, computer vision, and time series forecasting, with a focus on scalable model deployment and performance evaluation. Passionate about applying AI to solve complex business problems and deliver production-ready solutions.

EDUCATION

Education Level	Institution	Board/Stream	Year	Score Type	Score
B.Tech	Atria University, Bengaluru	Computer Science (Data Science & AI)	2025 (3rd year)	CGPA	9.7
Class XII	Vidyagyan School, Bulandshahr	CBSE (PCM)	2021 - 2022	Percentage	94%
Class X	Vidyagyan School, Bulandshahr	CBSE	2019 - 2020	Percentage	96.33%

EXPERIENCE

AI Research Intern – Centre For Air Borne System (CABS), DRDO (Ministry of defence India) August 2025 - Present

- Developing a fully offline, secure RAG chatbot for rapid access to DRDO CABS manuals, enabling low-latency, context-aware retrieval.
- Implemented embeddings using MiniLM and vector storage with ChromaDB, generating responses via Llama 3 (8B) model.
- Designed scalable retrieval pipelines and optimized query response times for operational deployment.
- Conducted statistical evaluation of retrieval accuracy to ensure reliability for real-world users.

AI Research Intern – Climate & Sustainability Tech, ACAD (**link**) May 2025 - July 2025

- Built geospatial ML tools and web apps using Python, FastAPI, Leaflet.js, and Google Earth Engine for ecosystem monitoring.
- Processed satellite imagery (Sentinel, Landsat, MODIS), applied indices (NDVI, chlorophyll, soil moisture), and detected stubble-burning hotspots for Delhi-NCR.
- Developed interactive dashboards enabling real-time environmental analysis and decision-making.

Data Science Intern at Insignia Consultancy Solutions June 2024 - Dec 2024

- Developed Named Entity Recognition (NER) models for resume-job matching and conducted data annotation & preprocessing for supervised ML pipelines.
- Researched Generative AI models and optimized ML algorithms for enhanced text analysis.
- Contributed to feature engineering and model deployment workflows, improving inference speed and accuracy.

Trainee at Intel® Unnati Industrial Training Program - 2024 May 2024 - Jul 2024

- Under this training program I was given a project “Intel® Products Sentiment Analysis from Online Reviews”
- Developed a Python-based automated web scraper to efficiently extract relevant data from Amazon.com.
- Utilized Natural Language Processing (NLP) techniques to analyze sentiment in customer reviews and extract valuable insights.
- Built a machine learning model achieving an accuracy of 85% in predicting sentiment across different review categories.

PROJECTS

The Monk AI – RAG Chatbot| September 2025

- Built an advanced Retrieval-Augmented Generation (RAG) chatbot with dual interaction modes (Beginner & Expert).
- Implemented transformers, cross-encoder re-ranking, Hindi translation, and context-grounded responses with citations.
- Designed pipeline for low-latency, voice-enabled queries, similar to real-time recommendation/search use cases in fashion-tech.

AI Research Assistant (MCP Protocol) | August 2025 :

- Engineered a multi-agent research automation system using FastAPI, Socket.IO, and Gemini 1.5 Flash LLM.
- Designed modular agents (Orchestrator, Search, Summarizer, Report Writer) with real-time collaboration over MCP protocol.
- Integrated Google Search API for dynamic retrieval and built an interactive JS frontend for visualization.
- Optimized for real-time search and response, aligning with scalable search and recommendation pipelines.

Sentiment Analysis on Ear Wearable Product Reviews Using ML & DL (2023 - 2024)

- Built a deep learning–based sentiment analysis model (LSTM) achieving 90% accuracy and F1-score of 0.91 on 21,825 Amazon product reviews.
- Automated large-scale data collection via Selenium and implemented Flask-based REST API for real-time predictions with confidence scores.
- Applied supervised learning, text preprocessing, and hyperparameter optimization, enabling scalable deployment for e-commerce sentiment tracking.

Breast Tumor Classification (2023 - 2024)

- Developed a diagnostic ML/DL system analyzing 20,403 medical images across multiple phases (Non-Cancer, Early, Middle).
- Applied advanced preprocessing (Gaussian Smoothing, HOG feature extraction) and trained models including CNNs (80.6% accuracy), Random Forest (77%), and SVM (74%).
- Designed evaluation pipeline using precision, recall, and F1-score, ensuring clinical reliability.
- Demonstrated low-latency inference capability to support scalable deployment in real-world healthcare setups.

TECH SKILLS

- **Programming Languages**: Python, SQL
- **ML/DL Expertise:** Regression, Classification, Clustering, Decision Trees, Random Forest, CNNs, RNNs, LSTMs, Transformers, Graph Neural Networks (GNNs), Representation Learning, Optimization Methods, Bayesian Modeling, Time Series Forecasting, Search Systems, Generative AI (RAG, LLMs, Prompt Engineering) Convolutional Neural Networks (CNNs), RNNs, LSTM, RAGs, Digital Image Processing, Feature Extraction, Supervised & Unsupervised Learning, Time Series Forecasting, Transformers, LLMs, HuggingFace.
- **Developer Tools & Databases** : Jupyter Notebook, Google Colab, GitHub, VS Code, Altair AI Studio, Docker, MongoDB, MySQL, VectorDB, MCP
- **Data & Deployment:** Feature Engineering, Data Pipelines, ETL, Model Deployment (real-time & batch), API Development, Containerization (Docker), Model Monitoring & Evaluation (A/B Testing, Precision, Recall, F1, AUC)
- **Developer Tools:** Jupyter Notebook, Google Colab, GitHub, VS Code, Altair AI Studio, MCP Protocol
- **Databases:** SQL (MySQL), NoSQL (MongoDB, VectorDBs – ChromaDB, Pinecone)

CERTIFICATION & ACHIEVEMENTS

- Research paper Published at IEEE Space
- All India **1st** position in Data Science Contest by Altair India.
- Research paper Published in HTL Journal
- Secured 6th position (Consolation Prize) in the IYD Hackathon organized by Sitare University, Lucknow, for developing an AI-powered Valmiki Ramayan Fact Checker.
- Machine Learning Professional Certicate by Altair
- IBM Certificate for Machine Learning (Coursera)
- Machine Learning with Python – GreatLearning
- Python for Machine Learning – GreatLearning
- Python Basic Certificate (HackerRank)
- University of Michigan – Python (Coursera)
- ChatGPT for Data Analytics
- SQL Basic Certificate (HackerRank)
- SQL Advanced Certificate (HackerRank)